

Supplement 1: Descriptive statistics and descriptions of missing data for the two databases

Table of Contents

clinicaltrials.gov.....	2
Missing data (clinicaltrials.gov).....	2
Registration and dates (clinicaltrials.gov).....	3
Inclusions/Exclusions (clinicaltrials.gov)	3
Study characteristics (clinicaltrials.gov).....	5
Study status (clinicaltrials.gov)	7
Funding (clinicaltrials.gov)	8
Sample size (clinicaltrials.gov)	8
ANZCTR.....	10
Missing data (ANZCTR).....	10
Registration and dates (ANZCTR).....	11
Inclusions/Exclusions (ANZCTR)	11
Study characteristics (ANZCTR).....	13
Health conditions (ANZCTR)	16
Study status (ANZCTR)	17
Sponsors and funding (ANZCTR).....	17
Sample size (ANZCTR)	19

clinicaltrials.gov

There were 10,100 studies excluded. The total number of remaining studies is 272,988. The data were extracted from clinicaltrials.gov on 01-Feb-2021.

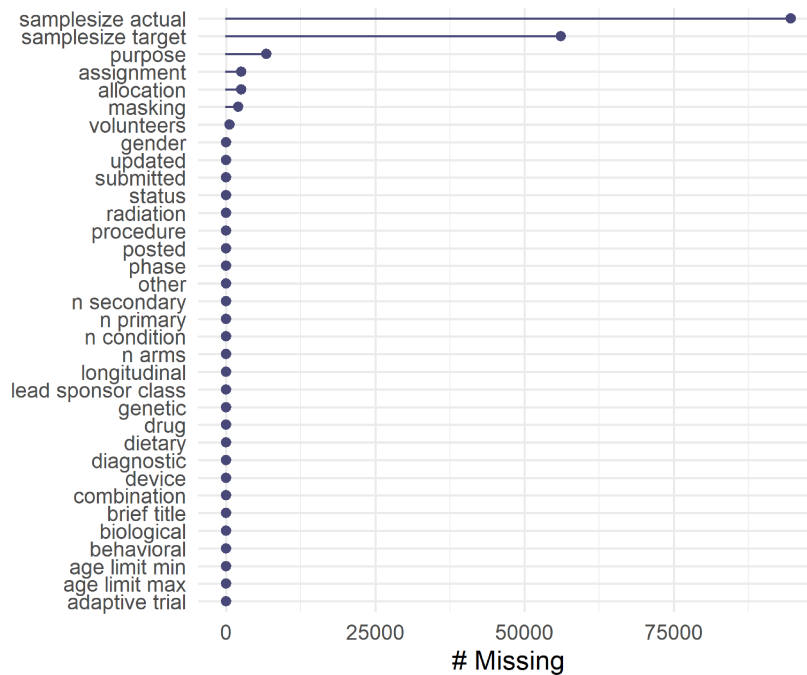
Missing data (clinicaltrials.gov)

The plots below summarise the missing data. The columns are variables and the rows are observations. The grey areas show missing data. The coloured plots show a random sample of 10,000 studies as the data are too large to show every study.



Counts of missing data (clinicaltrials.gov)

The plots below shows the counts of missing data using all studies.



The total number of studies in 272,988.

Overall the amount of missing data was small. The most frequent missing variable was the study purpose which is the main objective of the intervention(s) being evaluated by the clinical trial. This was missing for 2% of studies.

The logical variables cannot be missing by design, for example 'longitudinal' was created by whether the word "longitudinal" was mentioned in the detailed study description or not.

Many variables were mandatory and so cannot be missing, e.g., study status.

Registration and dates (clinicaltrials.gov)

Registration dates (clinicaltrials.gov)

date	n	median	q1	q3
posted	272,988	Feb 2015	Nov 2010	May 2018
updated	272,988	Jun 2018	May 2015	Apr 2020

The summary statistics are the median and inter-quartile range.

Inclusions/Exclusions (clinicaltrials.gov)

Gender included (clinicaltrials.gov)

	Freq	% Valid	% Total
All	232734	85	85

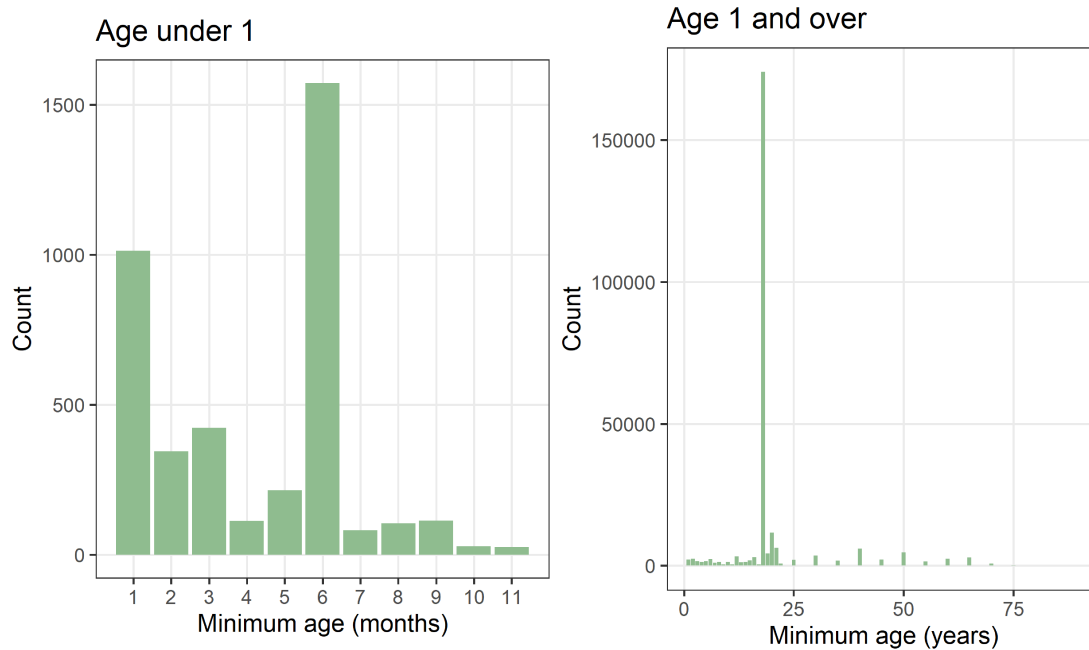
Female	26511	10	10
Male	13742	5	5
<NA>	1		0
Total	272988	100	100

Age included (clinicaltrials.gov)

Minimum age (clinicaltrials.gov)

	Freq	% Valid	% Total
Restricted	258396	95	95
No limit	14592	5	5
<NA>	0		0
Total	272988	100	100

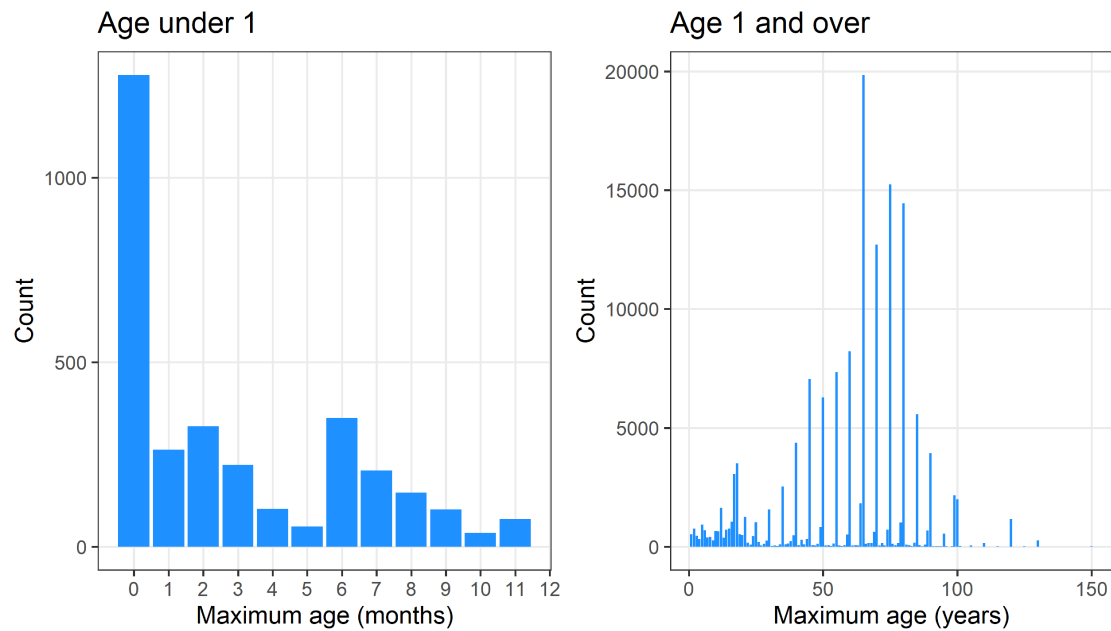
Minimum age limit in years (clinicaltrials.gov)



This plot is just for the 258,396 studies that had a minimum age limit. By far the most common minimum age is 18. The oldest exclusion for minimum age was 90 years.

Maximum age (clinicaltrials.gov)

	Freq	% Valid	% Total
Restricted	150717	55	55
No limit	122271	45	45
<NA>	0		0
Total	272988	100	100

Maximum age limit in years (clinicaltrials.gov)

This plot is just for the 150,717 studies that had a maximum age limit. The oldest exclusion for maximum age was 150 years.

Healthy volunteers included (clinicaltrials.gov)

	Freq	% Valid	% Total
No	202374	74	74
Yes	70079	26	26
<NA>	535		0
Total	272988	100	100

Study characteristics (clinicaltrials.gov)**Purpose (clinicaltrials.gov)**

	Freq	% Valid	% Total
Treatment	177513	67	65
Prevention	28945	11	11
Basic Science	13568	5	5
Supportive Care	12463	5	5
Diagnostic	12317	5	5
Other	12028	5	4
Health Services Research	6412	2	2
Screening	2261	1	1
Device Feasibility	754	0	0

<NA>	6727		2
Total	272988	100	100

Masking (clinicaltrials.gov)

	Freq	% Valid	% Total
None	152877	56	56
Single	36915	14	14
Double	35433	13	13
Quadruple	26844	10	10
Triple	18873	7	7
Missing	5	0	0
<NA>	2041		1
Total	272988	100	100

Assignment (clinicaltrials.gov)

	Freq	% Valid	% Total
Parallel	159496	59	58
Single Group	78591	29	29
Crossover	23781	9	9
Sequential	4582	2	2
Factorial	3995	1	1
Missing	7	0	0
<NA>	2536		1
Total	272988	100	100

Allocation (clinicaltrials.gov)

	Freq	% Valid	% Total
Randomized	178772	66	65
N/A	61995	23	23
Non-Randomized	29732	11	11
<NA>	2489		1
Total	272988	100	100

Phase (clinicaltrials.gov)

	Freq	% Valid	% Total
Early Phase 1	3464	1	1
N/A	120353	44	44
Phase 1	32331	12	12
Phase 1/Phase 2	11054	4	4
Phase 2	44189	16	16
Phase 2/Phase 3	5307	2	2
Phase 3	30120	11	11
Phase 4	26170	10	10

<NA>	0	0
Total	272988	100

Adaptive/platform trial (clinicaltrials.gov)

	Freq	%
FALSE	272820	100
TRUE	168	0
Total	272988	100

Was an adaptive or platform trial mentioned in the study description.

Longitudinal (clinicaltrials.gov)

	Freq	%
FALSE	270702	99
TRUE	2286	1
Total	272988	100

Was longitudinal mentioned in the study description.

Number of outcomes (clinicaltrials.gov)

	n_primary	n_secondary
Min	0	0
Q1	1	1
Median	1	2
Q3	2	5
Max	214	459

Number of primary and secondary outcomes. A small number of studies had a huge number of outcomes.

Study status (clinicaltrials.gov)

	Freq	%
Completed	149721	55
Recruiting	39237	14
Unknown status	26671	10
Terminated	17641	6
Not yet recruiting	13659	5
Active, not recruiting	13370	5
Withdrawn	8019	3
Enrolling by invitation	2070	1
Suspended	1393	1
Unknown status [Previously: Active, not recruiting]	1195	0
Unknown status [Previously: Enrolling by invitation]	12	0
Total	272988	100

Funding (clinicaltrials.gov)

Lead sponsor type (clinicaltrials.gov)

	Freq	%
Other	190273	70
Industry	74612	27
NIH	4941	2
U.S. Fed	3162	1
Total	272988	100

Sample size (clinicaltrials.gov)

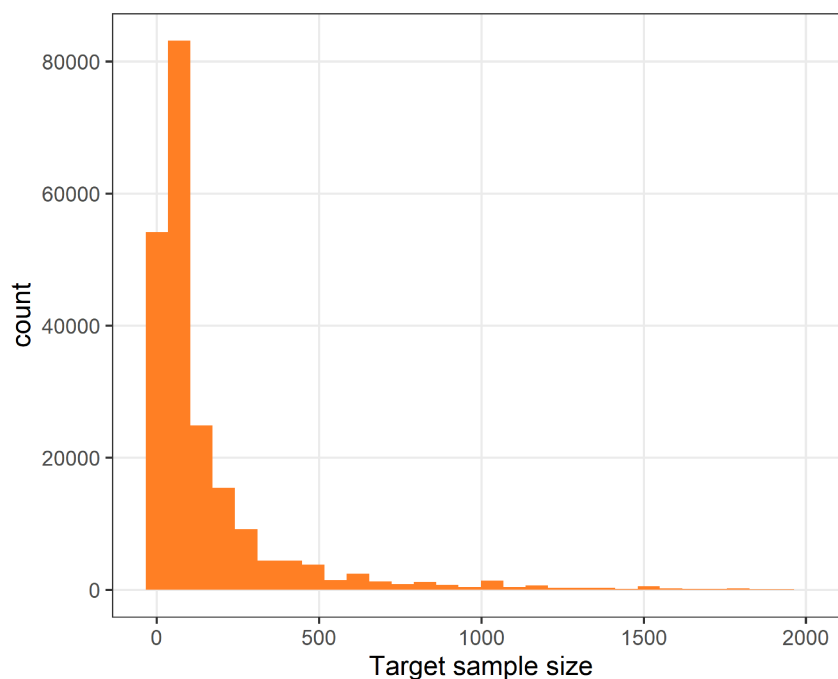
Target and achieved sample size (clinicaltrials.gov)

	Achieved	Target
Min	0	0
Q1	22	35
Median	52	70
Q3	139	178
Max	20121212	20000000

The table shows summary statistics.

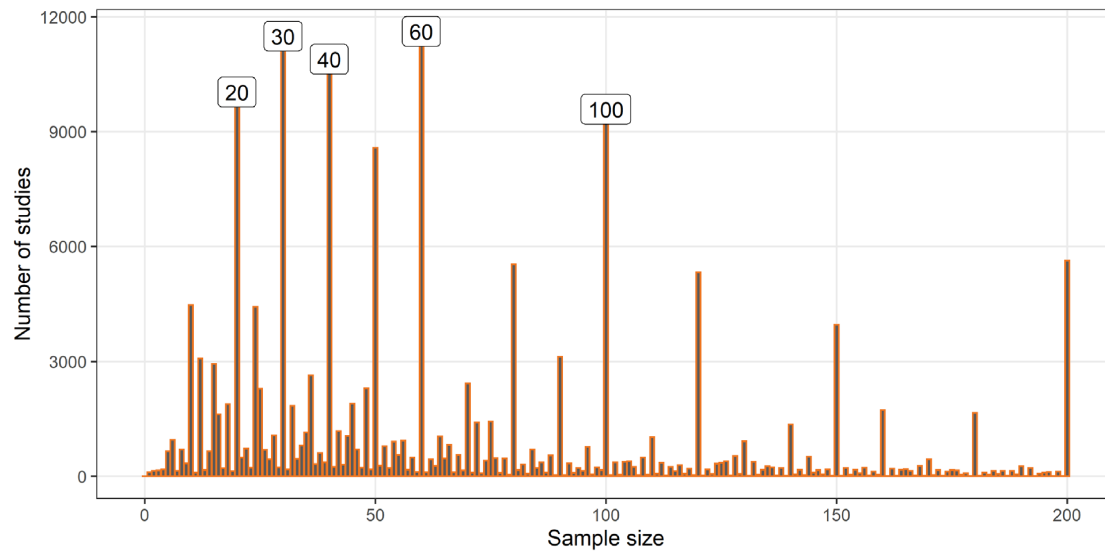
There were 5 studies with a zero target sample size.

Histogram of sample size (clinicaltrials.gov)



The plot above is for the 212,443 studies with a target sample size under 2,000.

Histogram looking for digit preference in target sample size (clinicaltrials.gov)



The plot above is for the 171,768 studies with a target sample size of 200 or fewer.

There is a strong digit preference at multiples of ten. The modal sample size is . The top five sample sizes are shown for each sample size type.

ANZCTR

These summary statistics are for trials registered in ANZCTR and not clinicaltrials.gov.

The total number of studies is 17,531. The data were extracted from ANZCTR on 01-Feb-2020.

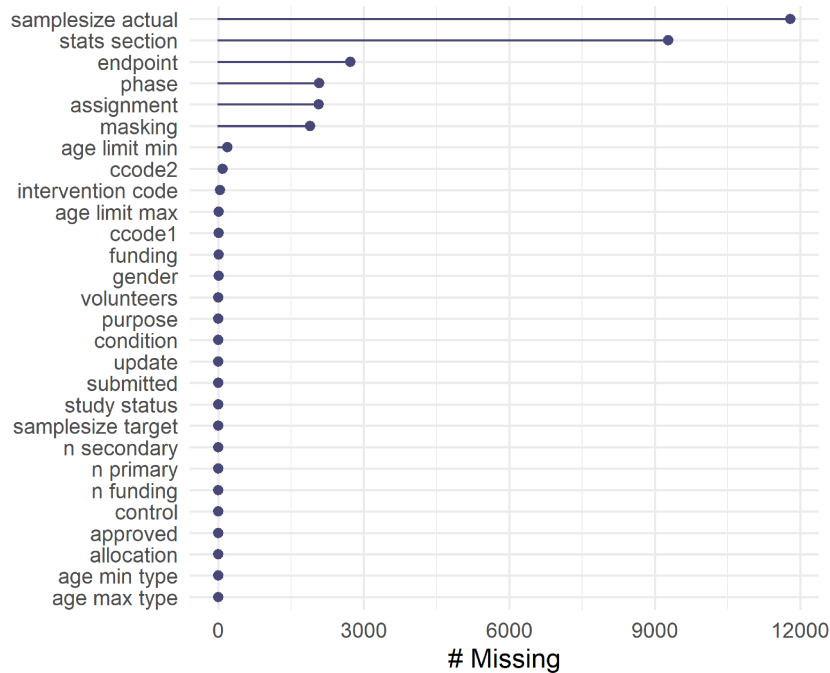
Missing data (ANZCTR)

The columns are variables and the rows are observations. The grey areas show missing data.



Counts of missing data (ANZCTR)

The plots below shows the counts of missing data.



Registration and dates (ANZCTR)

Registration dates (ANZCTR)

date	n	median	q1	q3
submitted	17,531	Oct 2014	Jan 2011	Dec 2017
approved	17,531	Nov 2014	Feb 2011	Jan 2018
update	17,531	Apr 2017	Apr 2013	Jun 2019

The summary statistics are the median and inter-quartile range.

Anticipated start date (ANZCTR)

min	q1	median	q3	max
Jun 1982	Jan 2010	Nov 2013	Oct 2017	Dec 2022

The summary statistics are the minimum, inter-quartile range, median and maximum.

Inclusions/Exclusions (ANZCTR)

Gender included (ANZCTR)

	Freq	% Valid	% Total
All	14900	85	85
Females	1701	10	10
Males	926	5	5

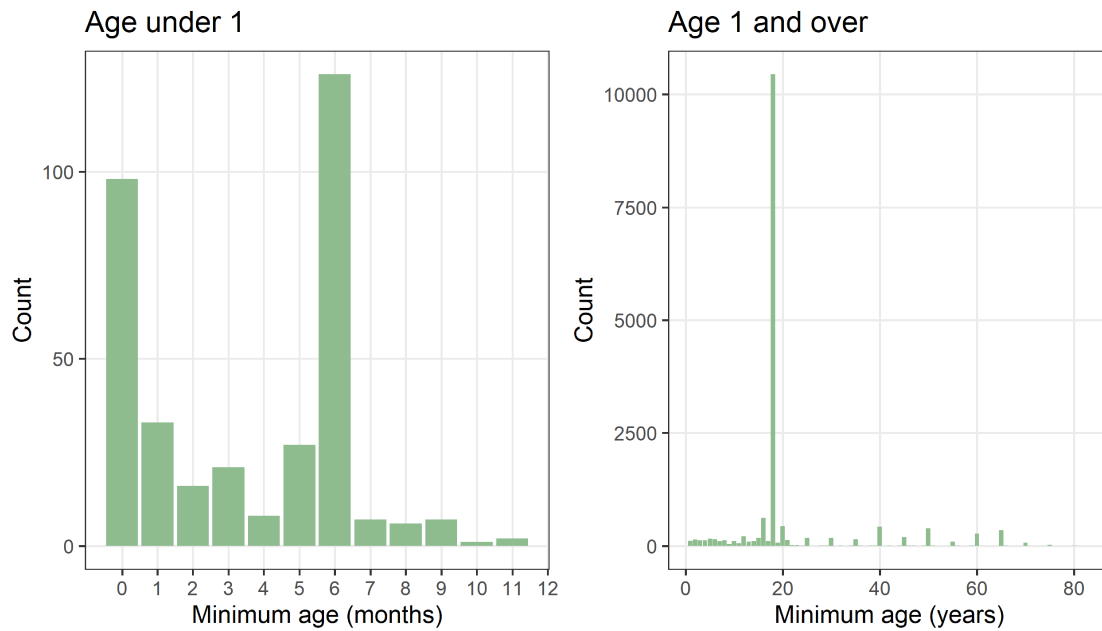
<NA>	4	0
Total	17531	100

Age included (ANZCTR)

Minimum age (ANZCTR)

	Freq	% Valid	% Total
Restricted	16623	95	95
No limit	632	4	4
Not stated	271	2	2
<NA>	5	0	0
Total	17531	100	100

Minimum age limit in years (ANZCTR)

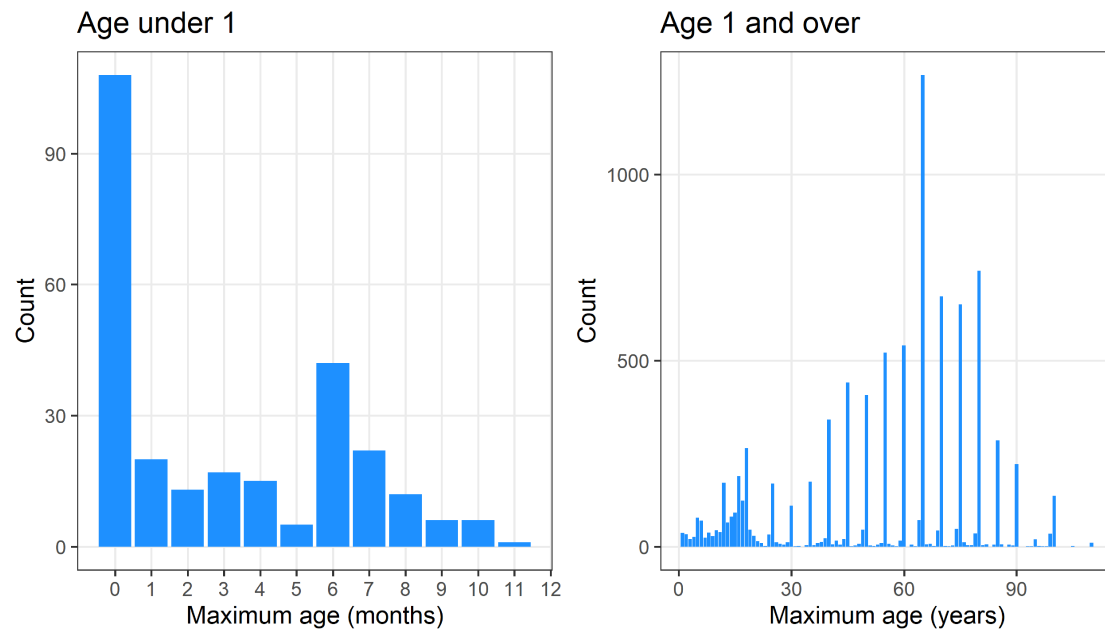


This plot is just for the 16,438 studies that had a minimum age limit. By far the most common minimum age is 18. The oldest exclusion for minimum age was 83 years.

Maximum age (ANZCTR)

	Freq	% Valid	% Total
Restricted	9092	52	52
No limit	7613	43	43
Not stated	821	5	5
<NA>	5	0	0
Total	17531	100	100

Maximum age limit in years (ANZCTR)



This plot is just for the 9,081 studies that had a maximum age limit. The oldest exclusion for maximum age was 110 years.

There is a strong digit preference in the age limit with spikes in multiples of five years.

Healthy volunteers included (ANZCTR)

	Freq	% Valid	% Total
No	12907	74	74
Yes	4621	26	26
<NA>	3		0
Total	17531	100	100

Study characteristics (ANZCTR)

Purpose (ANZCTR)

	Freq	% Valid	% Total
Treatment	12056	69	69
Prevention	3257	19	19
Educational / counselling / training	1432	8	8
Diagnosis	784	4	4
<NA>	2		0
Total	17531	100	100

Assignment (ANZCTR)

	Freq	% Valid	% Total
Parallel	10020	65	57
Single group	2427	16	14
Crossover	1925	12	11
Other	863	6	5
Factorial	226	1	1
<NA>	2070		12
Total	17531	100	100

Allocation (ANZCTR)

	Freq	% Valid	% Total
Randomised controlled trial	13155	75	75
Non-randomised trial	4376	25	25
<NA>	0		0
Total	17531	100	100

Masking (ANZCTR)

	Freq	% Valid	% Total
Blinded (masking used)	8274	53	47
Open (masking not used)	7366	47	42
<NA>	1891		11
Total	17531	100	100

Intervention code (ANZCTR)

	Freq	% Valid	% Total
Treatment: Drugs	4473	26	26
Treatment: Other	3502	20	20
Prevention	1868	11	11
Treatment: Devices	1802	10	10
Behaviour	1260	7	7
Rehabilitation	1237	7	7
Lifestyle	970	6	6
Other interventions	667	4	4
Treatment: Surgery	555	3	3
Diagnosis / Prognosis	514	3	3
Early detection / Screening	408	2	2
None	230	1	1
Not applicable	11	0	0
<NA>	34		0
Total	17531	100	100

Endpoint (ANZCTR)

	Freq	% Valid	% Total
Efficacy	9489	64	54
Safety/efficacy	3868	26	22
Safety	745	5	4
Pharmacokinetics	226	2	1
Bio-equivalence	152	1	1
Pharmacokinetics / pharmacodynamics	145	1	1
Pharmacodynamics	99	1	1
Bio-availability	86	1	0
<NA>	2721		16
Total	17531	100	100

Phase (ANZCTR)

	Freq	% Valid	% Total
Not Applicable	9830	64	56
Phase 0	58	0	0
Phase 1	1161	8	7
Phase 1 / Phase 2	331	2	2
Phase 2	1239	8	7
Phase 2 / Phase 3	269	2	2
Phase 3	907	6	5
Phase 3 / Phase 4	251	2	1
Phase 4	1401	9	8
<NA>	2084		12
Total	17531	100	100

Control (ANZCTR)

	Freq	% Valid	% Total
Active	10584	60	60
Placebo	3527	20	20
Uncontrolled	2780	16	16
Dose comparison	337	2	2
Historical	303	2	2
<NA>	0		0
Total	17531	100	100

Number of outcomes (ANZCTR)

	n_primary	n_secondary
Min	0	0
Q1	1	1
Median	1	3
Q3	2	6

Max 12 57

Number of primary and secondary outcomes.

Health conditions (ANZCTR)

Condition code #1 (ANZCTR)

	Freq	%
Mental Health	2029	12
Cancer	1592	9
Musculoskeletal	1157	7
Cardiovascular	1075	6
Respiratory	985	6
Public Health	978	6
Neurological	936	5
Anaesthesiology	927	5
Diet and Nutrition	917	5
Metabolic and Endocrine	906	5
Physical Medicine / Rehabilitation	671	4
Infection	664	4
Reproductive Health and Childbirth	611	3
Oral and Gastrointestinal	609	3
Surgery	478	3
Renal and Urogenital	380	2
Eye	352	2
Inflammatory and Immune System	345	2
Other	341	2
Injuries and Accidents	302	2
Skin	294	2
Stroke	283	2
Alternative and Complementary Medicine	262	1
Blood	211	1
Human Genetics and Inherited Disorders	95	1
Ear	85	0
Emergency medicine	38	0
Total	17523	100

Condition code #2, top 10 (ANZCTR)

	Freq	%
Other	12634	72
Other muscular and skeletal disorders	665	4
Diabetes	643	4
Depression	502	3
Health promotion/education	502	3
Obesity	475	3
Other diet and nutrition disorders	435	2

Other diseases of the mouth, teeth, oesophagus, digestive system including liver and colon	430	2
Anaesthetics	423	2
Other respiratory disorders / diseases	420	2
Physiotherapy	402	2
Total	17531	100

There are 155 condition codes, hence we just show the top ten. Conditions outside the top ten have been grouped together as 'other'.

Condition as free text, top 10 (ANZCTR)

	Freq	%
Other	15873	91
obesity	291	2
depression	263	2
stroke	204	1
type 2 diabetes	176	1
breast cancer	143	1
malaria	127	1
prostate cancer	125	1
asthma	114	1
anxiety	108	1
osteoarthritis	107	1
Total	17531	100

There are 8,395 conditions as free text, hence we just show the top ten.

This field is free text and so relies on what researchers write. Some categories could likely be combined, e.g., "prostate cancer" and "prostate cancer patients".

Study status (ANZCTR)

Study status (ANZCTR)

	Freq	%
Completed	7175	41
Not yet recruiting	4259	24
Recruiting	4072	23
Active, not recruiting	1042	6
Stopped early	602	3
Withdrawn	303	2
Suspended	78	0
Total	17531	100

Sponsors and funding (ANZCTR)

Top institutes (ANZCTR)

	Freq	%

Other	15163	86
University of Sydney	494	3
University of Queensland	305	2
Monash University	252	1
Royal Adelaide Hospital	244	1
Curtin University	231	1
University of Melbourne	204	1
University of South Australia	163	1
Deakin University	159	1
Griffith University	158	1
Royal Melbourne Hospital	158	1
Total	17531	100

There are 7,831 institutes, hence we just show the top ten.

Funding (ANZCTR)

Number of funders (ANZCTR)

	Freq	%
0	767	4
1	14046	80
2	2044	12
3	456	3
4	133	1
5	43	0
6	16	0
7	7	0
8	9	0
9	6	0
11	2	0
12	1	0
14	1	0
Total	17531	100

Funding type (ANZCTR)

	Freq	%
Government body	4570	22
University	3930	19
Commercial sector/Industry	3706	18
Charities/Societies/Foundations	3561	17
Hospital	2602	13
Self funded/Unfunded	922	4
Other Collaborative groups	646	3
Other	630	3
Total	20567	100

This includes multiple results per study.

Funders, top 10 (ANZCTR)

	Freq	%
Other	17879	87
NHMRC	1736	8
ARC	189	1
Health Research Council of New Zealand	170	1
Monash University	124	1
University of Otago	119	1
Curtin University	77	0
The University of Sydney	76	0
University of Sydney	70	0
Deakin University	65	0
Health Research Council	63	0
Total	20568	100

This includes multiple results per study.

There are 10,588 funders, hence we just show the top ten.

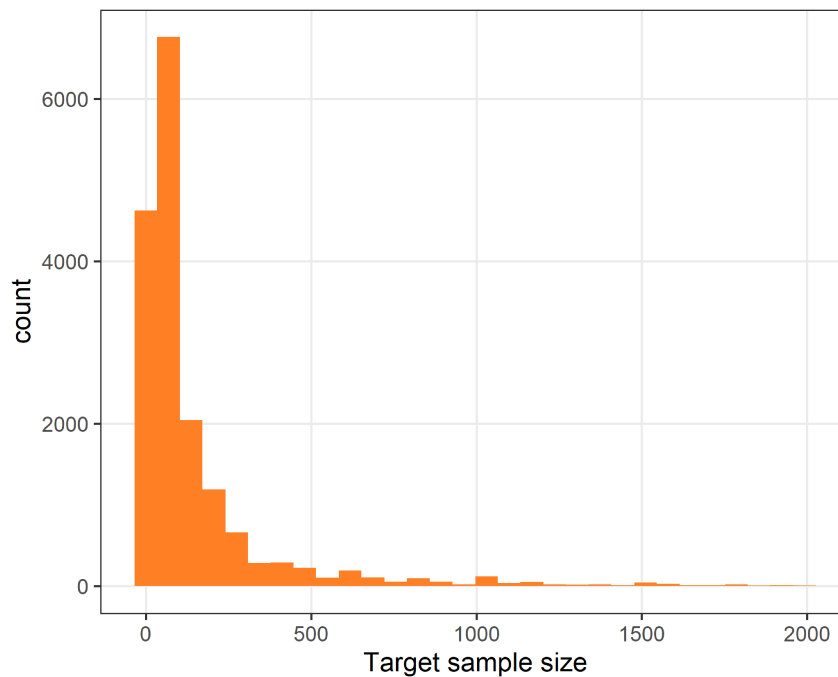
Sample size (ANZCTR)**Target and achieved sample size (ANZCTR)**

	Achieved	Target
Min	0	0
Q1	26	31
Median	60	66
Q3	140	158
Max	139977	714306

The table shows summary statistics.

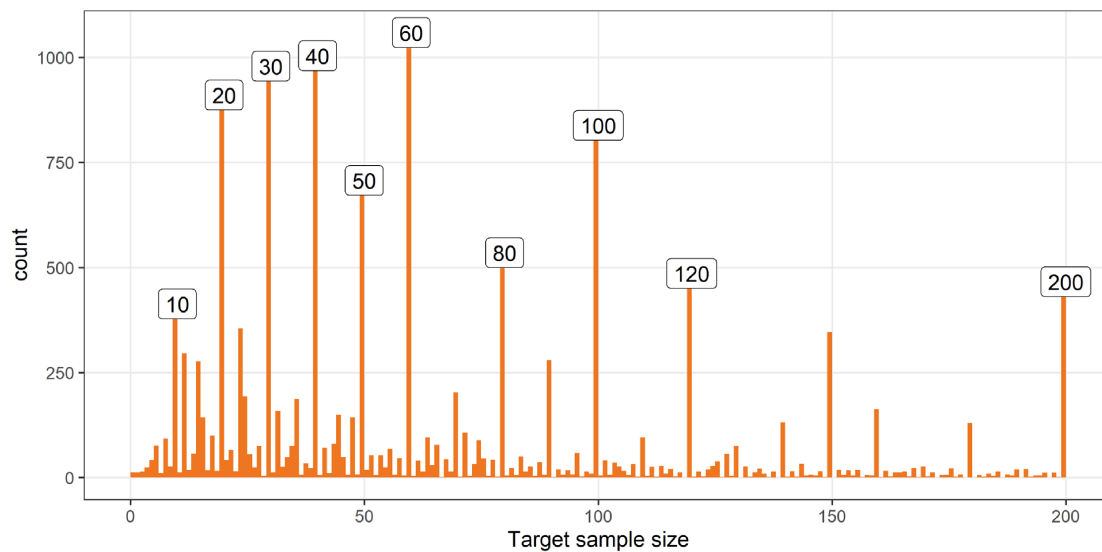
There were 2 studies with a zero target sample size.

Histogram of sample size (ANZCTR)



The plot above is for the 17,136 studies with a target sample size under 2,000.

Histogram looking for digit preference in target sample size (ANZCTR)



The plot above is for the 14,210 studies with a target sample size of 200 or fewer.

There is a strong digit preference at multiples of ten. The modal sample size is 60 and this is dominated by withdrawn studies.